Collective behavior of stock price movements in an emerging market

Raj Kumar Pan* and Sitabhra Sinha†
The Institute of Mathematical Sciences, C.I.T. Campus, Taramani, Chennai - 600 113 India
(Dated: December 3, 2008)

To investigate the universality of the structure of interactions in different markets, we analyze the cross-correlation matrix **C** of stock price fluctuations in the National Stock Exchange (NSE) of India. We find that this *emerging* market exhibits strong correlations in the movement of stock prices compared to *developed* markets, such as the New York Stock Exchange (NYSE). This is shown to be due to the dominant influence of a common market mode on the stock prices. By comparison, interactions between related stocks, e.g., those belonging to the same business sector, are much weaker. This lack of distinct sector identity in emerging markets is explicitly shown by reconstructing the network of mutually interacting stocks. Spectral analysis of **C** for NSE reveals that, the few largest eigenvalues deviate from the bulk of the spectrum predicted by random matrix theory, but they are far fewer in number compared to, e.g., NYSE. We show this to be due to the relative weakness of intra-sector interactions between stocks, compared to the market mode, by modeling stock price dynamics with a two-factor model. Our results suggest that the emergence of an internal structure comprising multiple groups of strongly coupled components is a signature of market development.

PACS numbers: 89.65.Gh,05.45.Tp,89.75.-k

I. INTRODUCTION

Financial markets can be considered as complex systems having many interacting elements and exhibiting large fluctuations in their associated observable properties, such as stock price or market index [1, 2]. The state of the market is governed by interactions among its components, which can be either traders or stocks. In addition, market activity is also influenced significantly by the arrival of external information. Statistical properties of stock price fluctuations and correlations between price movements of different stocks have been analyzed by physicists in order to understand and model financial market dynamics [3, 4]. The fluctuation distribution of stock prices is found to follow a power law with exponent $\alpha \sim 3$, the so-called "inverse cubic law" [5, 6]. This property is quite robust, and has been seen in developed as well as emerging markets [7]. On the other hand, it is not yet known whether the cross-correlation behavior between stock price fluctuations has a similar universal nature. Although the existence of collective modes have been inferred from the study of market dynamics, such studies have almost exclusively focussed on developed markets, in particular, the New York Stock Exchange (NYSE).

To uncover the structure of interactions among the elements in a financial market, physicists primarily focus on the spectral properties of the correlation matrix of stock price movements. Pioneering studies investigated whether the properties of the empirical correlation matrix differed from those of a random matrix that would have been obtained had the price movements been uncorrelated [8, 9]. Such deviations from the predictions of random matrix theory (RMT) can provide clues about the underlying interactions between various stocks. It was observed that, while the bulk of the eigenvalue distribution for the correlation matrix of NYSE and Tokyo Stock Exchange follow the spectrum predicted by RMT [8, 9, 10, 11], the few largest eigenvalues deviate significantly from this. The largest eigenvalue has been identified as representing the influence of the entire market, common for all stocks, whereas, the remaining large eigenvalues are associated with the different business sectors, as indicated by the composition of their corresponding eigenvectors [10, 12]. The interaction structure of stocks in NYSE have been reconstructed using filtering techniques implementing matrix decomposition [13] or maximum likelihood clustering [14]. Correlation matrix analysis has applications in the area of financial risk management, as mutually correlated price movements may indicate the presence of strong interactions between stocks [15]. Such analyses have been performed using asset trees and asset graphs to obtain the taxonomy of an optimal portfolio of stocks [16, 17, 18].

While it is generally believed that stock prices in emerging markets tend to be relatively more correlated than the developed ones [19], there have been very few studies of the former in terms of analysing the spectral properties of correlation matrices [20, 21, 22, 23]. Most studies of correlated price movements in emerging markets have looked at the *synchronicity* which measures the incidence of similar (i.e., up or down) price movements across stocks [19, 24]. Although related to correlation the two measures are not same, as correlation also gives the relative magnitude of similarity. In this paper, we analyze the cross-correlations among stocks in the Indian financial market, one of the largest emerging markets in

*Electronic address: rajkp@imsc.res.in
†Electronic address: sitabhra@imsc.res.in

the world. Our study spans over 1996-2006, a period which coincides with the decade of rapid transformation of the Indian economy after liberalization in the early 1990s.

We find that, in terms of the properties of its collective modes, the Indian market shows significant deviations from developed markets. As the fluctuation distribution of stocks in the Indian market [7, 21, 25] follows the "inverse cubic law" seen in NYSE [6, 26], the deviations observed in the correlation properties should be almost entirely due to differences in the nature of interaction structure in the two markets. Our observation of a higher degree of correlation in the Indian market compared to developed markets is found to be the result of a dominant market mode affecting all the stocks, which is further accentuated by the relative absence of clusters of mutually interacting stocks. This is explicitly shown by reconstructing the network of interactions within the market, using a filtered correlation matrix from which the common market influence and random noise has been removed. This procedure give a more accurate representation of the intra-market structure than the commonly used method of constructing minimal spanning tree from the unfiltered correlation matrix [13, 16, 17]. To support the interpretation of our empirical observations, we present a two-factor model of market dynamics in Section IV. Multi-factor models of market behavior have been used by other groups for explaining various spectral properties of empirical correlation matrices [27, 28, 29]. In this model, we assume the market to consist of several correlated groups of stocks which are also influenced by a common external signal, i.e., market mode. By varying the relative strength of the factor associated with the market mode to that associated with the groups, we show that decreasing the intra-group interactions result in spectral distribution properties similar to that seen for the Indian market. Our results imply that one of the key features signifying the transition of a market from emerging to developed status is the appearance and consolidation of distinct group identities.

II. DATA ANALYZED

The National Stock Exchange (NSE) is the largest stock market in India. Having commenced operations from Nov 1994, it is already the world's third largest stock exchange (after NASDAQ and NYSE) in terms of transactions [30]. It is thus an excellent source of data for studying the correlation structure of price movements in an *emerging* market.

We have considered the daily closing price data of 201 stocks (see Table I) traded in NSE from Jan 1996 to May 2006, which corresponds to 2607 days. This data is obtained from the NSE web-site [31] and has been manually corrected by us for stock splitting. The selected stocks were traded over the entire period 1996-2006 and had the minimum number of missing data points (i.e., days for which no price data is available). If the price value of a stock is missing on a particular day, a problem common to data from emerging markets [20], it is assumed that no trading took place on that day, i.e, the price remained the same as the preceding day. For comparison we also consider the daily closing price of 434 stocks of NYSE belonging to the S&P 500 index over the same period as the Indian data. However, the total number of working days is slightly different, viz., 2622 days. This data was obtained from the Yahoo! Finance website [32]. In all our analysis, while comparing with the NSE data, we have used multiple random samples of 201 stocks each, from the set of 434 NYSE stocks. We verified that the results obtained were independent of the particular sample of 201 stocks chosen.

To ensure that the missing closing prices in the Indian market data do not result in artifacts leading to spurious divergence from the US market, we have also performed our analysis on synthetic US market data containing the same number of missing data points. Multiple sets of such data were generated from the actual closing price time series by randomly choosing the required number of data points and replacing them with the same value as the preceding day. The resulting analysis showed no significant difference from the results obtained with the original US data.

III. THE RETURN CROSS-CORRELATION MATRIX

To observe correlation between the price movements of different stocks, we first measure the price fluctuations such that the result is independent of the scale of measurement. If $P_i(t)$ is price of the stock $i = 1, \ldots, N$ at time t, then the (logarithmic) price return of the ith stock over a time interval Δt is defined as

$$R_i(t, \Delta t) \equiv \ln P_i(t + \Delta t) - \ln P_i(t). \qquad (1)$$

As different stocks have varying levels of volatility (measured by the standard deviation of its returns) we define the normalized return,

$$r_i(t, \Delta t) \equiv \frac{R_i - \langle R_i \rangle}{\sigma_i}, \qquad (2)$$

where $\sigma_i \equiv \sqrt{\langle R_i^2 \rangle - \langle R_i \rangle^2}$, is the standard deviation of R_i and $\langle \ldots \rangle$ represents time average over the period of observation. We then compute the equal time cross-correlation matrix $\mathbf{C}$, whose element

$$C_{ij} \equiv \langle r_i r_j \rangle, \qquad (3)$$

represents the correlation between returns for stocks i and j. By construction, $\mathbf{C}$ is symmetric with $C_{ii} = 1$ and C_{ij} has a value in the domain $[-1, 1]$. Fig. 1 shows that, the correlation among stocks in NSE is larger on the average compared to that among the stocks in NYSE.

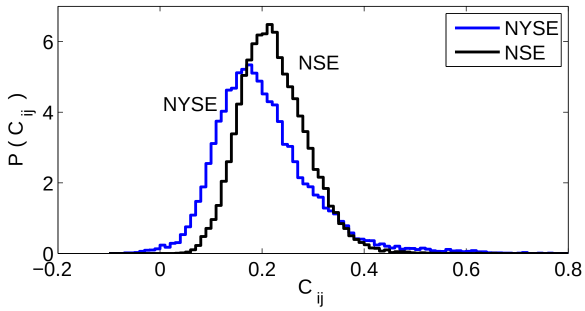

FIG. 1: The probability density function of the elements of the correlation matrix **C** for 201 stocks in the NSE of India and NYSE for the period Jan 1996-May 2006. The mean value of elements of **C** for NSE and NYSE, $\langle C_{ij} \rangle$, are 0.22 and 0.20 respectively.

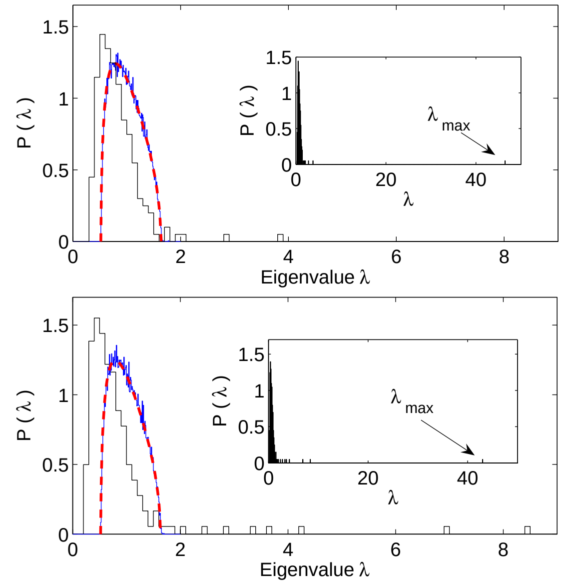

FIG. 2: The probability density function of the eigenvalues of the correlation matrix **C** for NSE (top) and NYSE (bottom). For comparison, the theoretical distribution predicted by Eq. (4) is shown using broken curves, which overlaps with the distribution obtained from the surrogate correlation matrix generated by randomly shuffling each time series. In both figures, the inset shows the largest eigenvalue.

This supports the general belief that developing markets tend to be more correlated than developed ones. To understand the reason behind this excess correlation, we perform an eigenvalue analysis of the correlation matrix.

A. Eigenvalue spectrum of correlation matrix

If the N return time series of length T are mutually uncorrelated, then the resulting random correlation matrix is called a Wishart matrix, whose statistical properties are well known [33]. In the limit $N \to \infty$, $T \to \infty$, such that $Q \equiv T/N \geq 1$, the eigenvalue distribution of this random correlation matrix is given by

$$P_{\rm rm}(\lambda) = \frac{Q}{2\pi} \frac{\sqrt{(\lambda_{max} - \lambda)(\lambda - \lambda_{min})}}{\lambda}, \qquad (4)$$

for $\lambda_{min} \leq \lambda \leq \lambda_{max}$ and, 0 otherwise. The bounds of the distribution are given by $\lambda_{max,min} = [1 \pm (1/\sqrt{Q})]^2$. We now compare this with the statistical properties of the empirical correlation matrix for the NSE. In the NSE data, there are $N = 201$ stocks each containing $T = 2606$ returns; as a result $Q = 12.97$. Therefore, it follows that, in the absence of any correlation among the stocks, the distribution should be bounded between $\lambda_{min} = 0.52$ and $\lambda_{max} = 1.63$. As observed in developed markets [8, 9, 10, 11], the bulk of the eigenvalue spectrum $P(\lambda)$ for the empirical correlation matrix is in agreement with the properties of a random correlation matrix spectrum $P_{\rm rm}(\lambda)$, but a few of the largest eigenvalues deviate significantly from the RMT bound (Fig. 2). However, the number of these deviating eigenvalues are relatively few for NSE compared to NYSE. To verify that these outliers are not an artifact of the finite length of the observation period, we have randomly shuffled the return time series for each stock, and then re-calculated the resulting correlation matrix. The eigenvalue distribution for this surrogate matrix matches exactly with the random matrix spectrum $P_{\rm rm}(\lambda)$, indicating that the outliers are not due to "measurement noise" but are genuine indicators of correlated movement among the stocks. Therefore, by analyzing the deviating eigenvalues, we may be able to obtain an understanding of the structure of interactions between the stocks in the market.

B. Properties of the "deviating" eigenvalues

The largest eigenvalue λ_0 for the NSE cross-correlation matrix is more than 28 times greater than the maximum predicted by RMT. This is comparable to NYSE, where λ_0 is about 26 times greater than the random matrix upper bound. Upon testing with synthetic US data containing same number of missing data points as in the Indian market, we observed that λ_0 remains almost unchanged compared to the value obtained from the original US data. The corresponding eigenvector shows a relatively uniform composition, with all stocks contributing to it and all elements having the same sign (Fig. 3, top). As this is indicative of a common factor that affects all the stocks with the same bias, the largest eigenvalue is associated with the *market mode*, i.e., the collective response of the entire market to external information [8, 10]. Of more interest for understanding the market structure are the intermediate eigenvalues, i.e., those occurring between the largest eigenvalue and the bulk of the distribution predicted by RMT. For the NYSE, it was shown that corresponding eigenvectors of these eigenvalues are localized, i.e., only a small number of stocks, belonging to

similar or related businesses, contribute significantly to each of these modes [10, 12]. However, for NSE, although the Technology and the IT & Telecom stocks are dominant contributors to the eigenvector corresponding to the third largest eigenvalue, a direct inspection of eigenvector composition does not yield a straightforward interpretation in terms of a related group of stocks corresponding to any particular eigenvalue (Fig. 3).

To obtain a quantitative measure of the number of stocks contributing to a given eigenmode, we calculate the inverse participation ratio (IPR), defined for the kth eigenvector as $I_k \equiv \sum_{i=1}^{N}[u_{ki}]^4$, where u_{ki} are the components of eigenvector k. An eigenvector having components with equal value, i.e., $u_{ki} = 1/\sqrt{N}$ for all i, has $I_k = 1/N$. We find this to be approximately true for the eigenvector corresponding to the largest eigenvalue, which represents the market mode. To see how different stocks contribute to the remaining eigenvectors, we note that if a single stock had a dominant contribution in any eigenvector, e.g., $u_{k1} = 1$ and $u_{ki} = 0$ for $i \neq 1$, then $I_k = 1$ for that eigenvector. Thus, IPR gives the reciprocal of the number of eigenvector components (and therefore, stocks) with significant contribution. On the other hand, the average value of I_k, for eigenvectors of a random correlation matrix obtained by randomly shuffling the time series of each stock, is $\langle I \rangle = 3/N \approx 1.49 \times 10^{-2}$. Fig. 4 shows that the eigenvalues belonging to the bulk of the spectrum indeed have this value of IPR. But at the lower and higher end of eigenvalues, both the US and Indian markets show deviations, suggesting the existence of localized modes. However, these deviations are much less significant and fewer in number in the latter compared to the former. This implies that distinct groups, whose members are mutually correlated in their price movement, do exist in NYSE, while their existence is far less clear in NSE.

C. Filtering the correlation matrix

The above analysis suggests the existence of a market-induced correlation across all stocks, which makes it difficult to observe the correlations that might be due to interactions between stocks belonging to the same sector. Therefore, we now use a filtering method to remove market mode, as well as the random noise [13]. The correlation matrix is first decomposed as

$$\mathbf{C} = \sum_{i=0}^{N-1} \lambda_i \mathbf{u}_i \mathbf{u}_i^T, \tag{5}$$

where λ_i are the eigenvalues of $\mathbf{C}$ sorted in descending order and $\mathbf{u}_i$ are corresponding eigenvectors. As only the eigenvectors corresponding to the few largest eigenvalues are believed to contain information on significantly correlated stock groups, the contribution of the intra-group correlations to the $\mathbf{C}$ matrix can be written as a partial sum of $\lambda_\alpha \mathbf{u}_\alpha \mathbf{u}_\alpha^T$, where α is the index of the corresponding eigenvalue. Thus, the correlation matrix can be decomposed into three parts, corresponding to the *market*, *group* and *random* components:

$$\begin{aligned} \mathbf{C} &= \mathbf{C}^{market} + \mathbf{C}^{group} + \mathbf{C}^{random} \\ &= \lambda_0 \mathbf{u}_0 \mathbf{u}_0^T + \sum_{i=1}^{N_g} \lambda_i \mathbf{u}_i \mathbf{u}_i^T + \sum_{i=N_g+1}^{N-1} \lambda_i \mathbf{u}_i \mathbf{u}_i^T, \end{aligned} \tag{6}$$

where, N_g is the number of eigenvalues (other than the largest one) which deviates from the bulk of the eigenvalue spectrum. For NSE we have chosen $N_g = 5$. However, the exact value of this choice is not crucial as small changes in N_g do not alter the results, the error involved being limited to the eigenvalues closest to the bulk that have the smallest contribution to $\mathbf{C}^{group}$. Fig. 5 shows the result of decomposing the correlation matrix into the three components, for both the Indian and US markets. Compared to the latter, the distribution of matrix elements of $\mathbf{C}^{group}$ in the former shows a significantly truncated tail. This indicates that intra-group correlations are not prominent in NSE, whereas they are comparable with the overall market correlations in NYSE. It follows that the collective behavior in the Indian market is dominated by external information that affects all stocks. Correspondingly, correlations generated by interactions between stocks, as would be the case for stocks in a given business sector, are much weaker, and hence, such correlated sectors would be difficult to observe.

We indeed find this to be true when we use the information in the group correlation matrix to construct the network of interacting stocks [13]. The adjacency matrix $\mathbf{A}$ of this network is generated from the group correlation matrix $\mathbf{C}^{group}$ by using a threshold c_{th} such that $A_{ij} = 1$ if $C_{ij}^{group} > c_{th}$, and $A_{ij} = 0$ otherwise. Thus, a pair of stocks are connected if the group correlation coefficient C_{ij}^{group} is larger than a preassigned threshold value, c_{th}. To determine an appropriate choice of $c_{th} = c^*$ we observe the number of isolated clusters (a cluster being defined as a group of connected nodes) in the network for a given c_{th} (Fig. 6). We found this number to be much less in NSE compared to that observed in NYSE for any value of c_{th} [13]. Fig. 7 shows the resultant network for $c^* = 0.09$, for which the largest number of isolated clusters of stocks are obtained. The network has 52 nodes and 298 links partitioned into 3 isolated clusters. From these clusters, only two business sectors can be properly identified, namely the Technology and the Pharmaceutical sectors. The fact that the majority of the NSE stocks cannot be arranged into well-segregated groups reflecting business sectors illustrates our conclusion that intra-group interaction is much weaker than the market-wide correlation in the Indian market.

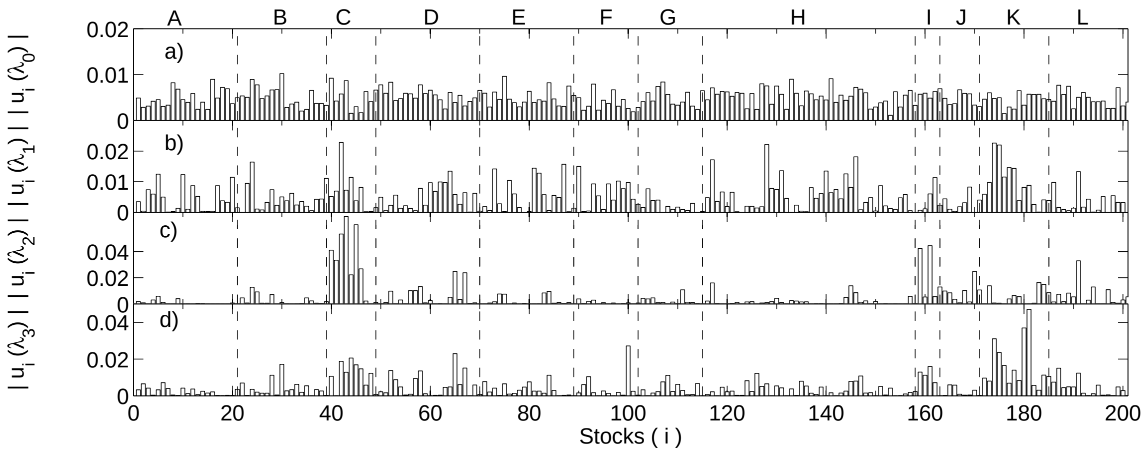

FIG. 3: The absolute values of the eigenvector components $u_i(\lambda)$ of stock i corresponding to the first four largest eigenvalues of $\mathbf{C}$ for NSE. The stocks i are arranged by business sectors separated by broken lines. A: Automobile & transport, B: Financial, C: Technology D: Energy, E: Basic materials, F: Consumer goods, G: Consumer discretionary, H: Industrial, I: IT & Telecom, J: Services, K: Healthcare & Pharmaceutical, L: Miscellaneous.

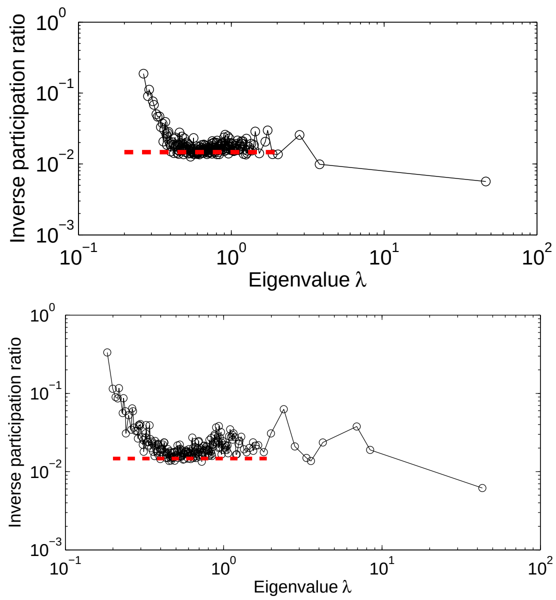

FIG. 4: Inverse participation ratio as a function of eigenvalue for the correlation matrix $\mathbf{C}$ of NSE (top) and NYSE (bottom). The broken line indicates the average value of $\langle I\rangle = 1.49\times10^{-2}$ for the eigenvectors of a matrix constructed by randomly shuffling each of the N time series.

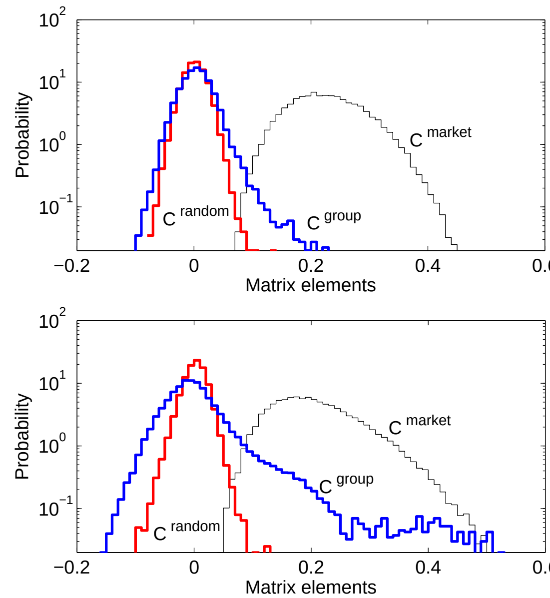

FIG. 5: The distribution of elements of correlation matrix corresponding to the market, $\mathbf{C}^{market}$, the group, $\mathbf{C}^{group}$, and the random interaction, $\mathbf{C}^{random}$. For NSE (top) $N_g = 5$ whereas for NYSE (bottom) $N_g = 10$. The short tail for the distribution of the $\mathbf{C}^{group}$ elements in NSE indicates that the correlation generated by mutual interaction among stocks is relatively weak.

D. Relating correlation with market evolution

We now compare between two different time intervals in the NSE data. For convenience we divide the data set into two non-overlapping parts corresponding to the periods between Jan 1996-Dec 2000 (Period I) and between Jan 2001-May 2006 (Period II). The corresponding correlation matrices $\mathbf{C}$ are generated following the same set of steps as for the entire data set. The average value for the elements of the correlation matrix is slightly lower for the later period, suggesting a greater homogeneity between the stocks at the earlier period (Fig. 8).

Next, we look at the eigenvalue distribution of $\mathbf{C}$ for the two periods (Fig. 9. The Q value for Period I is 6.21, while for Period II it is 6.77. Thus the bounds for the random distribution is almost same in the two cases. In contrast, the largest deviating eigenvalues, λ_0, are dif-

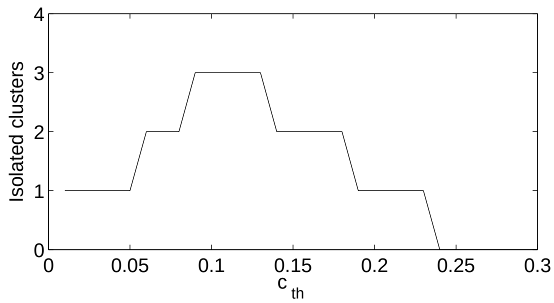

FIG. 6: The number of isolated clusters in the interaction network for NSE stocks as a function of the threshold value c_{th}. For low c_{th} the network consist of a single cluster which contains all the nodes, whereas for high c_{th} the network consists only of isolated nodes.

ferent: 48.56 for Period I and 45.88 for Period II. This implies the relative dominance of the market mode in the earlier period, again suggesting that with time the market has become less homogeneous. The number of deviating eigenvalues remain the same for the two periods.

When the interaction networks between stocks are generated for the two periods, they show less distinction into clearly defined sectors than was obtained with the data for the entire period. This is possibly because the shorter data sets create larger fluctuations in the correlation values, thereby making it difficult to segregate the existing market sectors. However, we do observe that, using the same threshold value for generating networks in the two periods yield, for the later period, isolated clusters that are distinguishable into distinct sub-clusters connected to each other via a few links only, whereas in the earlier period the clusters are much more homogeneous. This implies that as the Indian market is evolving, the interactions between stocks are tending to get arranged into clearly identifiable groups. We propose that such structural re-arrangement in the interactions is a hallmark of emerging markets as they evolve into developed ones.

IV. MODEL OF MARKET DYNAMICS

To understand the relation between the interaction structure among stocks and the eigenvalues of the correlation matrix, we perform a multivariate time series analysis using a simple two-factor model of market dynamics. We assume that the normalized return at time t of the ith stock from the kth business sector can be decomposed into (i) a market factor $r_m(t)$, that contains information or signal common to all stocks, (ii) a sector factor $r_g^k(t)$, representing effects exclusive to stocks in the kth sector, and (iii) an idiosyncratic term, $\eta_i(t)$, which corresponds to random variations unique for that stock. Thus,

$$r_i^k(t) = \beta_i r_m(t) + \gamma_i^k r_g^k(t) + \sigma_i \eta_i(t), \tag{7}$$

where β_i, γ_i^k and σ_i represent relative strengths of the three terms mentioned above, respectively. For simplicity, these strengths are assumed to be time independent. We choose $r_m(t)$, $r_g^k(t)$ and $\eta_i(t)$ from a zero mean and unit variance Gaussian distribution. We further assume that the normalized returns r_i, also follow Gaussian distribution with zero mean and unit variance. Although the empirically observed return distributions have power law tails, as these distributions are not Levy stable, they will converge to Gaussian if the returns are calculated over sufficiently long intervals. The assumption of unit variance for the returns ensures that the relative strengths of the three terms will follow the relation:

$${\beta_i}^2 + (\gamma_i^k)^2 + {\sigma_i}^2 = 1. \tag{8}$$

As a result, for each stock we can assign σ_i and γ_i independently, and obtain β_i from Eq. (8). We choose σ_i and γ_i from a uniform distribution having width δ and centered about the mean values σ and γ, respectively.

We now simulate an artificial market with N stocks belonging to K sectors by generating time series of length T for returns r_i^k from the above model. These K sectors are composed of $n_1, n_2, \ldots, n_K$ stocks such that $n_1 + n_2 + \cdots + n_K = N$. The collective behavior is then analysed by constructing the resultant correlation matrix $\mathbf{C}$ and obtaining its eigenvalues. Our aim is to relate the spectral properties of $\mathbf{C}$ with the underlying structure of the market given by the relative strength of the factors. We first consider the simple case, where the contribution due to market factor is neglected, i.e., $\beta_i = 0$ for all i, and the strength of sector factor is equal for all stocks within a sector, i.e., $\gamma_i^k = \gamma^k$, is independent of i. In this case, the spectrum of the correlation matrix is composed of K large eigenvalues, $1+(n_j-1)(\gamma^j)^2$, where $j = 1 \ldots K$, and $N-K$ small eigenvalues, $1-(\gamma^j)^2$, each with degeneracy $n_j - 1$, where $j = 1 \ldots K$ [28]. Now, we consider nonzero market factor which is equal for all stocks i.e., $\beta_i = \beta$ for all i, and the strength of sector factor is also same for all stocks, i.e., $\gamma_i^k = \gamma$ (independent of i and k). In this case too, there are K large eigenvalues and $N - K$ small eigenvalues. Our numerical simulations suggest that the largest and the second largest eigenvalues are

$$\begin{aligned} \lambda_0 &\sim N\beta^2, \\ \lambda_1 &\sim n_l(1-\beta^2), \end{aligned} \tag{9}$$

respectively, where n_l is the size of the largest sector, while the $N - K$ small degenerate eigenvalues are $1 - \beta^2 - \gamma^2$. We now choose the strength γ_i^k and σ_i from a uniform distribution with mean γ and σ respectively and with width $\delta = 0.05$. Fig. 10 shows the variation of the largest and second largest eigenvalues with σ and γ. The strength of the market factor is determined from Eq.8.

Note that, decreasing the strength of the sector factor relative to the market factor results in decreasing the second largest eigenvalue λ_1. As $Q = T/N$ is fixed, the RMT bounds for the bulk of the eigenvalue distribution, $[\lambda_{min}, \lambda_{max}]$, remain unchanged. Therefore, a decrease in

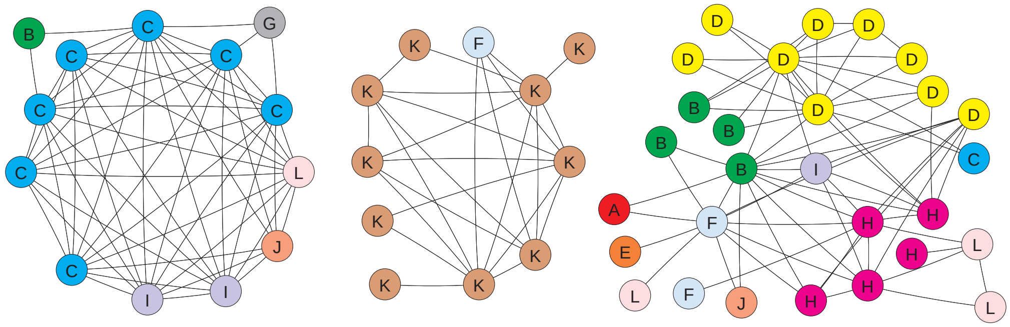

FIG. 7: The structure of interaction network in the Indian financial market at threshold $c^* = 0.09$. The left cluster comprises of mostly Technology stocks, while the middle cluster is composed almost entirely of Healthcare & Pharmaceutical stocks. By contrast, the cluster on the right is not dominated by any particular sector. The node labels indicate the business sector to which a stock belongs and are as specified in the caption to Fig 3.

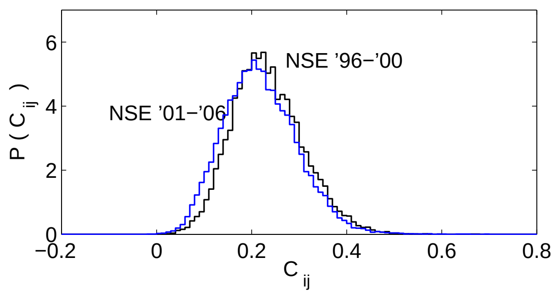

FIG. 8: The probability density functions of the elements in the correlation matrix **C** for NSE during (a) the period Jan 1996-Dec 2000 and (b) Jan 2001-May 2006. The mean value of the elements of **C** for the two periods are 0.23 and 0.21, respectively.

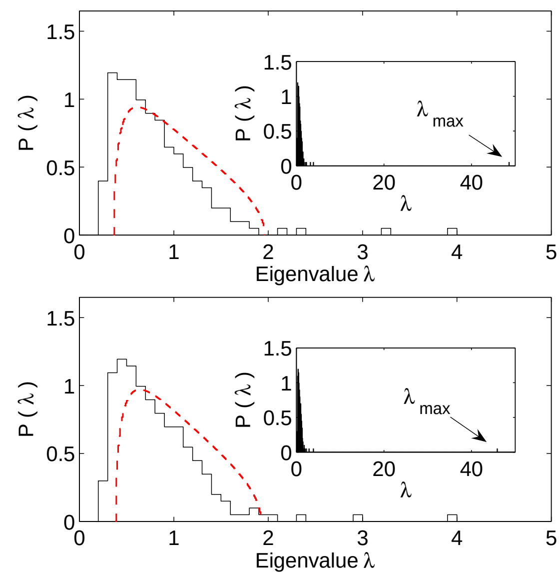

FIG. 9: The probability density function of the eigenvalues of the NSE correlation matrix **C** for the periods (top) Jan 1996-Dec 2000 and (bottom) Jan 2001-May 2006. For comparison, the theoretical distribution predicted by Eq. (4) is shown using broken curves. In both figures, the inset shows the largest eigenvalue.

λ_1 implies that the large intermediate eigenvalues occur closer to the bulk of the spectrum predicted by RMT, as is seen in the case of NSE. The analysis of the model supports our hypothesis that the spectral properties of the correlation matrix for the NSE are consistent with a market in which the effect of information common for all stocks (i.e., the market mode) is dominant, resulting in all stocks exhibiting a significant degree of correlation.

V. CONCLUSIONS

In conclusion, we demonstrate that the stocks in emerging market are much more correlated than in developed markets. Although, the bulk of the eigenvalue spectrum of the correlation matrix of stocks **C** in emerging market is similar to that observed for developed markets, the number of eigenvalues deviating from the RMT upper bound are smaller in number. Further, most of the observed correlation among stocks is found to be due to effects common to the entire market, whereas correlation due to interaction between stocks belonging to the same business sector are weak. This dominance of the market mode relative to modes arising through interactions between stocks makes an emerging market appear more correlated than developed markets. Using a simple two-factor model we show that a dominant market factor, relative to the sector factor, results in spectral properties similar to that observed empirically for the Indian market. Our study helps in understanding the evolution of markets as complex systems, suggesting that strong interactions may emerge within groups of stocks as a market evolves over time. How such self-organization occurs and its relation to other changes that a market undergoes during its development, e.g., large increase in transaction volume, is a question worth pursuing in the future with the tools available to econophysicists.

Our paper also makes a significant point regarding the

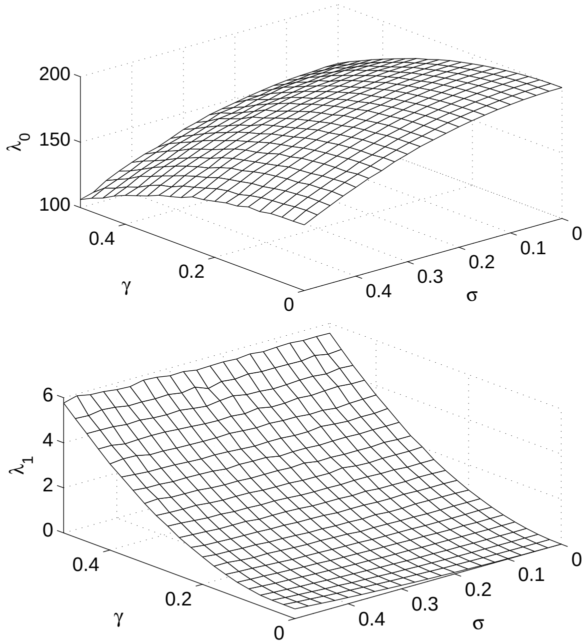

FIG. 10: The variation of the largest (top) and second largest (bottom) eigenvalues of the correlation matrix of simulated return in the two-factor model (Eq. 7) with the model parameters γ and σ (corresponding to strength of the sector and idiosyncratic effects, respectively). The matrix is constructed for $N = 200$ stocks each with return time series of length $T = 2000$ days. We assume there to be 10 sectors, each having 20 stocks.

physical understanding of markets as complex dynamical systems. In recent times, the role of the interaction structure within a market in governing its overall dynamical properties has come under increasing scrutiny. However, such intra-market interactions affect only very weakly certain market properties, which is underlined by the observation of identical fluctuation behaviour in markets having very different interaction structures, viz., NYSE and NSE [7, 25]. The system can be considered as a single homogeneous entity responding only to external signals in explaining these statistical features, e.g., the price fluctuation distribution. This suggests that the basic assumption behind the earlier approach of studying financial markets as essentially executing random walks in response to independent external shocks [34], which ignored the internal structure, may still be considered to be accurate for explaining market fluctuation phenomena. In other words, complex interacting systems like financial markets can have simple mean field-like description for some of their properties.

Acknowledgments

We thank N. Vishwanathan for assistance in preparing the data for analysis. We also thank M. Marsili and M.S. Santhanam for helpful discussions.

[1] R. N. Mantegna and H. E. Stanley, *Introduction to Econophysics* (Cambridge University Press, Cambridge, 1999).
[2] J. P. Bouchaud and M. Potters, *Theory of Financial Risk and Derivative Pricing* (Cambridge University Press, Cambridge, 2003), 2nd ed.
[3] I. Kondor and J. Kertesz, eds., *Econophysics: An Emerging Science* (Kluwer, Dordrecht, 1999).
[4] A. Chatterjee and B. K. Chakrabarti, eds., *Econophysics of Stock and other Markets* (Springer, Milan, 2006).
[5] T. Lux, Applied Financial Economics **6**, 463 (1996).
[6] V. Plerou, P. Gopikrishnan, L. A. Nunes Amaral, M. Meyer, and H. E. Stanley, Phys. Rev. E **60**, 6519 (1999).
[7] R. K. Pan and S. Sinha, Europhys. Lett. **77**, 58004 (2007).
[8] L. Laloux, P. Cizeau, J. P. Bouchaud, and M. Potters, Phys. Rev. Lett. **83**, 1467 (1999).
[9] V. Plerou, P. Gopikrishnan, B. Rosenow, L. A. Nunes Amaral, and H. E. Stanley, Phys. Rev. Lett. **83**, 1471 (1999).
[10] V. Plerou, P. Gopikrishnan, B. Rosenow, L. A. Nunes Amaral, T. Guhr, and H. E. Stanley, Phys. Rev. E **65**, 066126 (2002).
[11] A. Utsugi, K. Ino, and M. Oshikawa, Phys. Rev. E **70**, 026110 (2004).
[12] P. Gopikrishnan, B. Rosenow, V. Plerou, and H. E. Stanley, Phys. Rev. E **64**, 035106(R) (2001).
[13] D. H. Kim and H. Jeong, Phys. Rev. E **72**, 046133 (2005).
[14] L. Giada and M. Marsili, Phys. Rev. E **63**, 061101 (2001).
[15] H. M. Markowitz, *Portfolio Selection: : Efficient Diversification of Investments* (John Wiely & Sons, Inc., New York, 1959).
[16] R. N. Mantegna, Eur. Phys. Jour. B **11**, 193 (1999).
[17] J. P. Onnela, A. Chakraborti, K. Kaski, and J. Kertesz, Eur. Phys. Jour. B **30**, 285 (2002).
[18] J. P. Onnela, A. Chakraborti, K. Kaski, J. Kertesz, and A. Kanto, Phys. Rev. E **68**, 056110 (2003).
[19] R. Morck, B. Yeung, and W. Yu, Journal of Financial Economics **58**, 215 (2000).
[20] D. Wilcox and T. Gebbie, Physica A **375**, 584 (2007).
[21] S. Sinha and R. K. Pan, *Econophysics of Stock and Other Markets* (Springer, Milan, 2006), chap. The power (Law) of Indian markets: Analysing NSE and BSE trading statistics, pp. 24–34.
[22] V. Kulkarni and N. Deo, *Econophysics of Stock and Other Markets* (Springer, Milan, 2006), chap. A random matrix approach to volatility in an Indian financial market, pp. 35–48.
[23] S. Cukur, M. Eryigit, and R. Eryigit, Physica A **376**, 555 (2007).
[24] A. Durnev, K. Li, R. Morck, and B. Yeung, The Economics of Transition **12**, 593 (2004).
[25] R. K. Pan and S. Sinha, physics/0607014 (2006).

[26] P. Gopikrishnan, V. Plerou, L. A. Nunes Amaral, M. Meyer, and H. E. Stanley, Phys. Rev. E **60**, 5305 (1999).
[27] J. D. Noh, Phys. Rev. E **61**, 5981 (2000).
[28] F. Lillo and R. N. Mantegna, Phys. Rev. E **72**, 016219 (2005).
[29] H. E. Roman, M. Albergante, M. Colombo, F. Croccolo, F. Marini, and C. Riccardi, Phys. Rev. E **73**, 036129 (2006).
[30] Tech. Rep., National Stock Exchange (2004).
[31] http://www.nseindia.com/.
[32] http://finance.yahoo.com/.
[33] A. M. Sengupta and P. P. Mitra, Phys. Rev. E **60**, 3389 (1999).
[34] L. Bachelier, Annales Scientifiques de l'École Normale Supérieure Sér **3**, 21 (1900).

TABLE I: The list of 201 stocks of NSE analyzed in this paper.

i	Company	Sector	i	Company	Sector	i	Company	Sector
1	UCALFUEL	Automobiles Transport	68	IBP	Energy	135	HIMATSEIDE	Industrial
2	MICO	Automobiles Transport	69	ESSAROIL	Energy	136	BOMDYEING	Industrial
3	SHANTIGEAR	Automobiles Transport	70	VESUVIUS	Energy	137	NAHAREXP	Industrial
4	LUMAXIND	Automobiles Transport	71	NOCIL	Basic Materials	138	MAHAVIRSPG	Industrial
5	BAJAJAUTO	Automobiles Transport	72	GOODLASNER	Basic Materials	139	MARALOVER	Industrial
6	HEROHONDA	Automobiles Transport	73	SPIC	Basic Materials	140	GARDENSILK	Industrial
7	MAHSCOOTER	Automobiles Transport	74	TIRUMALCHM	Basic Materials	141	NAHARSPG	Industrial
8	ESCORTS	Automobiles Transport	75	TATACHEM	Basic Materials	142	SRF	Industrial
9	ASHOKLEY	Automobiles Transport	76	GHCL	Basic Materials	143	CENTENKA	Industrial
10	M&M	Automobiles Transport	77	GUJALKALI	Basic Materials	144	GUJAMBCEM	Industrial
11	EICHERMOT	Automobiles Transport	78	PIDILITIND	Basic Materials	145	GRASIM	Industrial
12	HINDMOTOR	Automobiles Transport	79	FOSECOIND	Basic Materials	146	ACC	Industrial
13	PUNJABTRAC	Automobiles Transport	80	BASF	Basic Materials	147	INDIACEM	Industrial
14	SWARAJMAZD	Automobiles Transport	81	NIPPONDENR	Basic Materials	148	MADRASCEM	Industrial
15	SWARAJENG	Automobiles Transport	82	LLOYDSTEEL	Basic Materials	149	UNITECH	Industrial
16	LML	Automobiles Transport	83	HINDALC0	Basic Materials	150	HINDSANIT	Industrial
17	VARUNSHIP	Automobiles Transport	84	SAIL	Basic Materials	151	MYSORECEM	Industrial
18	APOLLOTYRE	Automobiles Transport	85	TATAMETALI	Basic Materials	152	HINDCONS	Industrial
19	CEAT	Automobiles Transport	86	MAHSEAMLES	Basic Materials	153	CARBORUNIV	Industrial
20	GOETZEIND	Automobiles Transport	87	SURYAROSNI	Basic Materials	154	SUPREMEIND	Industrial
21	MRF	Automobiles Transport	88	BILT	Basic Materials	155	RUCHISOYA	Industrial
22	IDBI	Financial	89	TNPL	Basic Materials	156	BHARATFORG	Industrial
23	HDFCBANK	Financial	90	ITC	Consumer Goods	157	GESHIPPING	Industrial
24	SBIN	Financial	91	VSTIND	Consumer Goods	158	SUNDRMFAST	Industrial
25	ORIENTBANK	Financial	92	GODFRYPHLP	Consumer Goods	159	SHYAMTELE	Telecom
26	KARURVYSYA	Financial	93	TATATEA	Consumer Goods	160	ITI	Telecom
27	LAKSHVILAS	Financial	94	HARRMALAYA	Consumer Goods	161	HIMACHLFUT	Telecom
28	IFCI	Financial	95	BALRAMCHIN	Consumer Goods	162	MTNL	Telecom
29	BANKRAJAS	Financial	96	RAJSREESUG	Consumer Goods	163	BIRLAERIC	Telecom
30	RELCAPITAL	Financial	97	KAKATCEM	Consumer Goods	164	INDHOTEL	Services
31	CHOLAINV	Financial	98	SAKHTISUG	Consumer Goods	165	EIHOTEL	Services
32	FIRSTLEASE	Financial	99	DHAMPURSUG	Consumer Goods	166	ASIANHOTEL	Services
33	BAJAUTOFIN	Financial	100	BRITANNIA	Consumer Goods	167	HOTELEELA	Services
34	SUNDARMFIN	Financial	101	SATNAMOVER	Consumer Goods	168	FLEX	Services
35	HDFC	Financial	102	INDSHAVING	Consumer Goods	169	ESSELPACK	Services
36	LICHSGFIN	Financial	103	MIRCELECTR	Consumer Discretonary	170	MAX	Services
37	CANFINHOME	Financial	104	SURAJDIAMN	Consumer Discretonary	171	COSMOFILMS	Services
38	GICHSGFIN	Financial	105	SAMTEL	Consumer Discretonary	172	DABUR	Health Care
39	TFCILTD	Financial	106	VDOCONAPPL	Consumer Discretonary	173	COLGATE	Health Care
40	TATAELXSI	Technology	107	VDOCONINTL	Consumer Discretonary	174	GLAXO	Health Care
41	MOSERBAER	Technology	108	INGERRAND	Consumer Discretonary	175	DRREDDY	Health Care
42	SATYAMCOMP	Technology	109	ELGIEQUIP	Consumer Discretonary	176	CIPLA	Health Care
43	ROLTA	Technology	110	KSBPUMPS	Consumer Discretonary	177	RANBAXY	Health Care
44	INFOSYSTCH	Technology	111	NIRMA	Consumer Discretonary	178	SUNPHARMA	Health Care
45	MASTEK	Technology	112	VOLTAS	Consumer Discretonary	179	IPCALAB	Health Care
46	WIPRO	Technology	113	KECINTL	Consumer Discretonary	180	PFIZER	Health Care
47	BEML	Technology	114	TUBEINVEST	Consumer Discretonary	181	EMERCK	Health Care
48	ALFALAVAL	Technology	115	TITAN	Consumer Discretonary	182	NICOLASPIR	Health Care
49	RIIL	Technology	116	ABB	Industrial	183	SHASUNCHEM	Health Care
50	GIPCL	Energy	117	BHEL	Industrial	184	AUROPHARMA	Health Care
51	CESC	Energy	118	THERMAX	Industrial	185	NATCOPHARM	Health Care
52	TATAPOWER	Energy	119	SIEMENS	Industrial	186	HINDLEVER	Miscellaneous
53	GUJRATGAS	Energy	120	CROMPGREAV	Industrial	187	CENTURYTEX	Miscellaneous
54	GUJFLUORO	Energy	121	HEG	Industrial	188	EIDPARRY	Miscellaneous
55	HINDOILEXP	Energy	122	ESABINDIA	Industrial	189	KESORAMIND	Miscellaneous
56	ONGC	Energy	123	BATAINDIA	Industrial	190	ADANIEXPO	Miscellaneous
57	COCHINREFN	Energy	124	ASIANPAINT	Industrial	191	ZEETELE	Miscellaneous
58	IPCL	Energy	125	ICI	Industrial	192	FINCABLES	Miscellaneous
59	FINPIPE	Energy	126	BERGEPAINT	Industrial	193	RAMANEWSPR	Miscellaneous
60	TNPETRO	Energy	127	GNFC	Industrial	194	APOLLOHOSP	Miscellaneous
61	SUPPETRO	Energy	128	NAGARFERT	Industrial	195	THOMASCOOK	Miscellaneous
62	DCW	Energy	129	DEEPAKFERT	Industrial	196	POLYPLEX	Miscellaneous
63	CHEMPLAST	Energy	130	GSFC	Industrial	197	BLUEDART	Miscellaneous
64	RELIANCE	Energy	131	ZUARIAGRO	Industrial	198	GTCIND	Miscellaneous
65	HINDPETRO	Energy	132	GODAVRFERT	Industrial	199	TATAVASHIS	Miscellaneous
66	BONGAIREFN	Energy	133	ARVINDMILL	Industrial	200	CRISIL	Miscellaneous
67	BPCL	Energy	134	RAYMOND	Industrial	201	INDRAYON	Miscellaneous